The correlation structure of local cortical networks intrinsically results from recurrent dynamics

Moritz Helias[1], **Tom Tetzlaff**[1], **Markus Diesmann**[1,2]

[1]Inst. of Neuroscience and Medicine (INM-6) and Inst. for Advanced Simulation (IAS-6), Jülich Research Centre and JARA, Jülich, Germany

[2]Medical Faculty, RWTH Aachen University, Aachen, Germany

Correspondence to: Dr. Moritz Helias

tel: +49-2461-61-9467
m.helias@fz-juelich.de

Abstract

Correlated neuronal activity is a natural consequence of network connectivity and shared inputs to pairs of neurons, but the task-dependent modulation of correlations in relation to behavior also hints at a functional role. Correlations influence the gain of postsynaptic neurons, the amount of information encoded in the population activity and decoded by readout neurons and synaptic plasticity. Further, it affects the power and spatial reach of extracellular signals like the local-field potential. A theory of correlated neuronal activity accounting for recurrent connectivity as well as fluctuating external sources is currently lacking. In particular, it is unclear how the recently found mechanism of active decorrelation by negative feedback on the population level affects the network response to externally applied correlated stimuli. Here, we present such an extension of the theory of correlations in stochastic binary networks. We show that (1) for homogeneous external input, the structure of correlations is mainly determined by the local recurrent connectivity, (2) homogeneous external inputs provide an additive, unspecific contribution to the correlations, (3) inhibitory feedback effectively decorrelates neuronal activity, even if neurons receive identical external inputs, and (4) identical synaptic input statistics to excitatory and to inhibitory cells increases intrinsically generated fluctuations and pairwise correlations. We further demonstrate how the accuracy of mean-field predictions can be improved by self-consistently including correlations. As a byproduct, we show that the cancellation of correlations between the summed inputs to pairs of neurons does not originate from the fast tracking of external input, but from the suppression of fluctuations on the population level by the local network. The suppression of fluctuations on the population level is a necessary constraint, but not sufficient to determine the structure of correlations. Therefore, the structure of correlations does not follow from the fast tracking of external inputs.

Author summary

The co-occurrence of action potentials of pairs of neurons within short time intervals is known since long. Such synchronous events can appear time-locked to the behavior of an animal and also theoretical considerations argue for a functional role of synchrony. Early theoretical work tried to explain correlated activity by neurons transmitting common fluctuations due to shared inputs. This, however, overestimates correlations. Recently the recurrent connectivity of cortical networks was shown responsible for the observed low baseline correlations. Two different explanations were given: One argues that excitatory and inhibitory population activities closely follow the external inputs to the network, so that their effects on a pair of cells mutually cancel. Another explanation relies on negative recurrent feedback to suppress fluctuations in the population activity, equivalent to small correlations. In a biological neuronal network one expects both, external inputs and recurrence, to affect correlated activity. The present work extends the theoretical framework of correlations to include both contributions and explains their qualitative differences. Moreover the study shows that the arguments of fast tracking and recurrent feedback are not equivalent, only the latter correctly predicts the cell-type specific correlations.

1 Introduction

The spatio-temporal structure and magnitude of correlations in cortical neural activity have since long been subject of research for a variety of reasons: The experimentally observed task-dependent modulation of correlations points at a potential functional role. In the motor cortex of behaving monkeys, for example, synchronous action potentials appear at behaviorally relevant time points [1]. The degree of synchrony is modulated by task performance, and the precise timing of synchronous events follows a change of the behavioral paradigm after a phase of re-learning. In primary visual cortex, saccades (eye movements) are followed by brief periods of synchronized neural firing [2, 3]. Further, correlations and fluctuations depend on the attentive state of the animal [4], with higher correlations and slow fluctuations observed during quiet wakefulness, and faster, uncorrelated fluctuations in the active state [5]. It is still unclear whether the observed modulation of correlations is in fact employed by the brain, or whether it is merely an epiphenomenon. Theoretical studies have suggested a number of interpretations and mechanisms of

how correlated firing could be exploited: Correlations in afferent spike-train ensembles may provide a gating mechanism by modulating the gain of postsynaptic cells (for a review, see [6]). Synchrony in afferent spikes (or, more generally, synchrony in spike arrival) can enhance the reliability of postsynaptic responses and, hence, may serve as a mechanism for a reliable activation and propagation of precise spatio-temporal spike patterns [7, 8, 9, 10]. Further, it has been argued that synchronous firing could be employed to combine elementary representations into larger percepts [11, 12, 7, 13, 14]. While correlated firing may constitute the substrate for some en- and decoding schemes, it can be highly disadvantageous for others: The number of response patterns which can be triggered by a given afferent spike-train ensemble becomes maximal if these spike trains are uncorrelated [15]. In addition, correlations in the ensemble impair the ability of readout neurons to decode information reliably in the presence of noise (see e.g. [16, 15, 17]). Recent studies have indeed shown that biological neural networks implement a number of mechanisms which can efficiently decorrelate neural activity, such as the nonlinearity of spike generation [18], synaptic variability and failure [19, 20], short-term synaptic depression [20], and the recurrent network dynamics [21, 22, 17]. To study the significance of experimentally observed task-dependent correlations, it is essential to provide adequate null hypotheses: Which level and structure of correlations is to be expected in the absence of any task-related stimulus or behavior? Even in the simplest network models without time varying input, correlations in the neural activity emerge as a consequence of shared input [23, 24, 25] and recurrent connectivity [22, 26, 17, 27, 28]. Irrespective of the functional aspect, the spatio-temporal structure and magnitude of correlations between spike trains or membrane potentials carry valuable information about the properties of the underlying network generating these signals [24, 26, 29, 27, 28] and could therefore help constraining models of cortical networks. Further, the quantification of spike-train correlations is a prerequisite to understand how correlation sensitive synaptic plasticity rules, such as spike-timing dependent plasticity [30], interact with the recurrent network dynamics [31]. Finally, knowledge of the expected level of correlations between synaptic inputs is crucial for the correct interpretation of extracellular signals like the local-field potential (LFP) [32].

Previous theoretical studies on correlations in local cortical networks provide analytical expressions for the magnitude [25, 22, 17] and the temporal shape [33, 34, 27, 28] of average pairwise correlations, capture the influence of the connectivity on correlations [35, 36, 26, 29, 27, 37], and connect oscillatory network states emerging from delayed negative feedback [38] to the shape of correlation functions [28]. We have in particular shown recently that negative feedback loops, abundant in cortical networks, constitute an efficient decorrelation mechanism and therefore allow neurons to fire nearly independently despite substantial shared presynaptic input [17] (see also [35, 22, 39]). We further pointed out that in networks of excitatory (E) and inhibitory (I) neurons, the correlations between neurons of different cell type (EE, EI, II) differ in both magnitude and temporal shape, even if excitatory and inhibitory neurons have identical properties and input statistics [17, 28]. It remains unclear, however, how this cell-type specificity of correlations is affected by the connectivity of the network.

The majority of previous theoretical studies on cortical circuits is restricted to local networks driven by external sources representing thalamo-cortical or cortico-cortical inputs (e.g. [40, 41, 42]). Most of these studies emphasize the role of the local network connectivity (e.g. [43]). Despite the fact that inputs from remote (external) areas constitute a substantial fraction of all excitatory inputs (about 50% [7], see also [44, 45]), their spatio-temporal structure is often abstracted by assuming that neurons in the local network are independently driven by external sources. A priori, this assumption can hardly be justified: Neurons belonging to the local cortical network receive, at least to some extent, inputs from identical or overlapping remote areas, for example due to patchy (clustered) horizontal connectivity [46, 47]. Hence, shared-input correlations are likely to play a role not only for local but also for external inputs. Coherent activation of neurons in remote presynaptic areas constitutes another source of correlated external input, in particular for sensory areas [48, 5, 49, 4]. So far, it is largely unknown how correlated external input affects the dynamics of local cortical networks and alters correlations in their neural activity.

In this article, we investigate how the magnitude and the cell-type specificity of correlations depend on i) the connectivity in local cortical networks of finite size and ii) the level of correlations in external

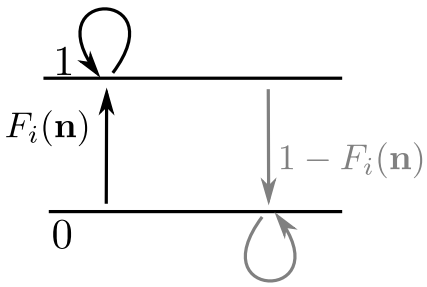

Figure 1: State transitions of a binary neuron. Each neuron is updated at random time points, intervals are i.i.d. exponential with mean duration τ, so the rate of updates per neuron i is τ^{-1}. The probability of neuron i to end in the up-state (1) is determined by the gain function $F_i(\mathbf{n})$ which potentially depends on the states $\mathbf{n}$ of all neurons in the network. The up-transitions are indicated by black arrows. The probability for the down state (0) is given by the complementary probability $1 - F_i(\mathbf{n})$, indicated by gray arrows.

inputs. Existing theories of correlations in cortical networks are not sufficient to address these questions as they either do not incorporate correlated external input [33, 17, 27, 26, 29] or assume infinitely large networks [22]. Lindner et al. [35] studied the responses of finite populations of spiking neurons receiving correlated external input, but described inhibitory feedback by a global compound process.

Our work builds on the existing theory of correlations in stochastic binary networks [33], a well-established model in the neuroscientific community [50, 22]. This model has the advantage of requiring for its analytical treatment elementary mathematical methods only. We employ the same network structure used in the work by Renart et al. [22] which relates the mechanism of recurrent decorrelation to the fast tracking of external signals (see [51] for a recent review). This choice enables us to reconsider the explanation of decorrelation by negative feedback [17], originally shown for networks of leaky integrate-and-fire neurons, and to compare it to the findings of Renart et al. In fact, the motivation for the choice of the model arose from the review process of [17], during which both the reviewers and the editors encouraged us to elucidate the relation of our work to the one of Renart et al. in a separate subsequent manuscript. The present work delivers this comparison.

The remainder of the manuscript is organized as follows: In **"Methods"**, we develop the theory of correlations in recurrent random networks of excitatory and inhibitory cells driven by fluctuating input from an external population of finite size. We present a method accounting for the fluctuations in the synaptic input to each cell, which effectively linearize the hard threshold of the neurons. We further include the resulting finite-size correlations into the established mean-field description [50, 52] to increase the accuracy of the theory. In **"Results"**, we first show in **"Correlations are driven by intrinsic and external fluctuations"** that correlations in recurrent networks are not only caused by the externally imposed correlated input, but also by intrinsically generated fluctuations of the local populations. We demonstrate that the external drive causes an overall shift of the correlations, but that their relative magnitude is mainly determined by the intrinsically generated fluctuations. In **"Cancellation of input correlations"**, we revisit the earlier reported phenomenon of the suppression of correlations between input currents to pairs of cells [22] and show that it is a direct consequence of the suppression of fluctuations on the population level [17]. Subsequently, in **"Influence of connectivity on the correlation structure"**, we investigate in how far the reported structure of correlations is a generic feature of balanced networks and isolate parameters of the connectivity determining this structure. Finally, in **"Discussion"**, we summarize our results and their implications for the interpretation of experimental data, discuss the limitations of the theory, and provide an outlook of how the improved theory may serve as a further building block to understand processing of correlated activity.

2 Methods

2.1 Networks of binary neurons. The state of a binary neuron is either 0 or 1, where 1 indicates activity, 0 inactivity [33, 53, 22]. The model shows stochastic transitions (at random points in time) between these two states controlled by transition probabilities, as illustrated in Figure 1. Using asynchronous update [54], in each infinitesimal interval $[t, t+\delta t)$ each neuron in the network has the probability $\frac{1}{\tau}\delta t$ to be chosen for update [55], where τ is the time constant of the neuronal dynamics. An equivalent implementation draws the time points of update independently for all neurons. For a particular neuron, the sequence of update points has exponentially distributed intervals with mean duration τ, i.e. update times form a Poisson process with rate τ^{-1}. We employ the latter implementation in the globally time-driven [56] spiking simulator NEST [57], and use a discrete time resolution $h = 0.1\,\mathrm{ms}$ for the intervals. The stochastic update constitutes a source of noise in the system. Given the i-th neuron is selected for update, the probability to end in the up-state ($n_i = 1$) is determined by the gain function $F_i(\mathbf{n})$ which possibly depends on the activity $\mathbf{n}$ of all other neurons. The probability to end in the down state ($n_i = 0$) is $1 - F_i(\mathbf{n})$. This model has been considered earlier [58, 33, 53], and here we follow the notation introduced in the latter work.

The state of the network of N such neurons is described by a binary vector $\mathbf{n} = (n_1, \ldots, n_N) \in \{0,1\}^N$. The stochastic system is completely characterized by the joint probability distribution $p(\mathbf{n})$ in all N binary variables $\mathbf{n}$. Knowing the joint probability distribution, arbitrary moments can be calculated, among them pairwise correlations. Here we are only concerned with the stationary state of the network. A stationary solution of $p(\mathbf{n})$ implies that for each state a balance condition holds, so that the incoming and outgoing probability fluxes sum up to zero. The occupation probability of the state is then constant. We denote as $\mathbf{n}_{i+} = (n_1, \ldots, n_{i-1}, 1, n_{i+1}, \ldots, n_N)$ the state, where the i-th neuron is active ($n_i = 1$), and $\mathbf{n}_{i-}$ where neuron i is inactive ($n_i = 0$). Since in each infinitesimal time interval at most one neuron can change state, for each given state $\mathbf{n}$ there are N possible transitions (each corresponding to one of the N neurons changing state). The sum of the probability fluxes into the state and out of the state must compensate to zero [59], so

$$0 = \tau \frac{\partial p(\mathbf{n})}{\partial t} = \sum_{i=1}^{N} \underbrace{(2n_i - 1)}_{\text{direction of flux}} \left(\underbrace{p(\mathbf{n}_{i-}) F_i(\mathbf{n}_{i-})}_{\text{neuron } i \text{ transition up}} - \underbrace{p(\mathbf{n}_{i+})(1 - F_i(\mathbf{n}_{i+}))}_{\text{neuron } i \text{ transition down}} \right) \quad \forall \quad \mathbf{n} \in \{0,1\}^N . \tag{1}$$

From this equation we derive expressions for the first $\langle n_k \rangle$ and second moments $\langle n_k n_l \rangle$ by multiplying with $n_k n_l$ and summing over all possible states $\mathbf{n} \in \{0,1\}^N$, which leads to

$$0 = \sum_{\mathbf{n} \in \{0,1\}^N} \sum_{i=1}^{N} n_k n_l (2n_i - 1) \underbrace{\left(p(\mathbf{n}_{i-}) F_i(\mathbf{n}_{i-}) - p(\mathbf{n}_{i+})(1 - F_i(\mathbf{n}_{i+})) \right)}_{\equiv G_i(\mathbf{n} \backslash n_i)} .$$

Note that the term denoted $G_i(\mathbf{n}\backslash n_i)$ does not depend on the state of neuron i. We use the notation $\mathbf{n}\backslash n_i$ for the state of the network excluding neuron i, i.e. $\mathbf{n}\backslash n_i = (n_1, \ldots, n_{i-1}, n_{i+1}, \ldots, n_N)$. Separating the terms in the sum over i into those with $i \neq k, l$ and the two terms with $i = k$ and $i = l$, we obtain

$$\begin{aligned} 0 &= \sum_{\mathbf{n}} \sum_{i=1, i\neq k,l}^{N} n_k n_l (2n_i - 1)\, G_i(\mathbf{n}\backslash n_i) + n_k n_l (2n_k - 1)\, G_k(\mathbf{n}\backslash n_k) + n_k n_l (2n_l - 1)\, G_l(\mathbf{n}\backslash n_l) \\ &= \sum_{i=1, i\neq k,l}^{N} \sum_{\mathbf{n}\backslash n_i} n_k n_l (G_i(\mathbf{n}\backslash n_i) - G_i(\mathbf{n}\backslash n_i)) + \sum_{\mathbf{n}} n_k n_l\, G_k(\mathbf{n}\backslash n_k) + \sum_{\mathbf{n}} n_k n_l\, G_l(\mathbf{n}\backslash n_l) \end{aligned}$$

where we obtained the first term by explicitly summing over state $n_i \in \{0,1\}$ (i.e. using $\sum_{\mathbf{n}\in\{0,1\}^N} = \sum_{\mathbf{n}\backslash n_i \in \{0,1\}^{N-1}} \sum_{n_i=0}^{1}$ and evaluating the sum $\sum_{n_1=0}^{1}$). This first sum obviously vanishes. The remaining terms are of identical form with the roles of k and l interchanged. We hence only consider the first of

them and obtain the other by symmetry. The first term simplifies to

$$\begin{aligned}
&\sum_{\mathbf{n}} n_k n_l \, G_k(\mathbf{n}\backslash n_k) \overset{n_k=1}{=} \sum_{\mathbf{n}\backslash n_k} n_l \, G_k(\mathbf{n}\backslash n_k) \\
\overset{\text{def. } G_k}{=} &\begin{cases} \sum_{\mathbf{n}\backslash n_k} p(\mathbf{n}_{k-}) \, F_k(\mathbf{n}_{k-}) + p(\mathbf{n}_{k+}) \, F_k(n_{k+}) - p(\mathbf{n}_{k+}) & \text{for } k = l \\ \sum_{\mathbf{n}\backslash n_k} p(\mathbf{n}_{k-}) \, n_l \, F_k(\mathbf{n}_{k-}) + p(\mathbf{n}_{k+}) \, n_l \, F_k(n_{k+}) - n_l \, p(\mathbf{n}_{k+}) & \text{for } k \neq l \end{cases} \\
= &\begin{cases} \langle F_k(\mathbf{n}) \rangle - \langle n_k \rangle & \text{for } k = l \\ \langle F_k(\mathbf{n}) \, n_l \rangle - \langle n_k n_l \rangle & \text{for } k \neq l \end{cases},
\end{aligned}$$

where we denote as $\langle f(\mathbf{n}) \rangle = \sum_{\mathbf{n} \in \{0,1\}^N} p(\mathbf{n}) f(\mathbf{n})$ the average of a function $f(\mathbf{n})$ with respect to the distribution $p(\mathbf{n})$. Taken together with the mirror term $k \leftrightarrow l$, we arrive at two conditions, one for the first ($k = l$, $\langle n_k^2 \rangle = \langle n_k \rangle$) and one for the second ($k \neq l$) moment

$$2\langle n_k n_l \rangle = \begin{cases} 2\langle F_k(\mathbf{n}) \rangle & \text{for } k = l \\ \langle F_k(\mathbf{n}) \, n_l \rangle + \langle F_l(\mathbf{n}) \, n_k \rangle & \text{for } k \neq l \end{cases}. \tag{2}$$

Considering the covariance $c_{kl} = \langle \delta n_k \delta n_l \rangle$ with centralized variables $\delta n_k = n_k - \langle n_k \rangle$, for $k \neq l$ one arrives at

$$2c_{kl} = \langle F_k(\mathbf{n}) \, \delta n_l \rangle + \langle F_l(\mathbf{n}) \, \delta n_k \rangle. \tag{3}$$

This equation is identical to eq. 3.9 in [33], to eqs. 3.12 and 3.13 in [53], and to eqs. (19)-(22) in [22, Supplementary material].

2.2 Mean-field solution. Starting from (1) for the general case $\frac{\partial p(\mathbf{n},t)}{\partial t} \neq 0$, a similar calculation as the one resulting in (2) for $k = l$ leads to

$$\tau \frac{\partial}{\partial t} \langle n_k \rangle = \langle F_k(\mathbf{n}) \rangle - \langle n_k \rangle,$$

where we used $\langle n_k^2 \rangle = \langle n_k \rangle$, valid for binary variables. As in [22] we now assume a particular form for the gain function and for the coupling between neurons by specifying

$$\begin{aligned}
F_k(\mathbf{n}) &= H\left(h_k - \theta\right) \\
h_k &= \sum_{l=1}^{N} J_{kl} n_l \\
H(x) &= \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases},
\end{aligned}$$

where J_{kl} is the incoming synaptic weight from neuron l to neuron k, H is the Heaviside function, and θ is the threshold of the activation function. We denote by h_k the summed synaptic input to the neuron, sometimes also called the "field". Because $n_k^2 = n_k$, the variance a_k of a binary variable is $a_k \equiv \langle n_k^2 \rangle - \langle n_k \rangle^2 = (1 - \langle n_k \rangle)\langle n_k \rangle$. We now aim to solve (2) for the case $k = l$, i.e. the equation $\langle n_k \rangle = \langle F_k \rangle$. In general, the right hand side depends on the fluctuations of all neurons projecting to neuron k. An exact solution is therefore complicated. However, for sufficiently irregular activity in the network we assume the neurons to be approximately independent. Further assume that in a network of homogeneous populations α (same parameters τ, θ and same statistics of the incoming connections for all neurons, i.e. same number $K_{\alpha\beta}$ and strength $J_{\alpha\beta}$ of incoming connections from neurons in a given population β) the mean activity of an individual neuron can be represented by the population mean

$m_\alpha = \langle \frac{1}{N_\alpha} \sum_{i\in\alpha} n_i \rangle$. The mean input to a neuron in population α then is

$$\langle h_\alpha \rangle \quad = \quad \sum_\beta K_{\alpha\beta} J_{\alpha\beta} m_\beta \equiv \mu_\alpha. \tag{4}$$

We assumed in the last step identical synaptic amplitudes $J_{\alpha\beta}$ for a synapse from a neuron in population β to a neuron in population α. So the input to each neuron has the same mean $\langle h_\alpha \rangle$. As a first approximation, if the mean activity in the network is not saturated, i.e. neither 0 nor 1, mapping this activity back by the inverse gain function to the input, h_α must be close to the threshold value, so

$$\langle h_\alpha \rangle \quad \simeq \quad \theta. \tag{5}$$

This relation may be solved for m_E and m_I to obtain a coarse estimate of the activity in the network [50, 52]. In mean-field approximation we assume that the fluctuations of the fields of individual neurons h_α around their mean are mutually independent, so that the fluctuations $\delta h_\alpha = h_\alpha - \langle h_\alpha \rangle$ of h_α are, in turn, caused by a sum of independent random variables and hence the variances add up to the variance σ_α^2 of the field

$$\langle \delta h_\alpha^2 \rangle \quad = \quad \sum_\beta K_{\alpha\beta} J_{\alpha\beta}^2 m_\beta (1 - m_\beta) \equiv \sigma_\alpha^2. \tag{6}$$

As h_α is a sum of typically thousands of synaptic inputs, it approaches a Gaussian distribution $h_\alpha \sim \mathcal{N}(\mu_\alpha, \sigma_\alpha^2)$ with mean μ_α and variance σ_α^2. In this approximation the mean activity in the network is the solution of

$$\begin{aligned} \tau \frac{\partial}{\partial t} m_\alpha + m_\alpha &= \langle F_\alpha(m_E, m_I, m_x) \rangle \quad \forall\, \alpha \in \{E, I\} \\ &\simeq \int_{-\infty}^{\infty} H(x - \theta)\, \mathcal{N}(\mu_\alpha, \sigma_\alpha^2, x)\, dx \\ &= \int_{\theta}^{\infty} \mathcal{N}(\mu_\alpha, \sigma_\alpha^2, x)\, dx \\ &= \frac{1}{2} \mathrm{erfc}\left(\frac{\theta - \mu_\alpha}{\sqrt{2}\sigma_\alpha} \right). \end{aligned} \tag{7}$$

This equation needs to be self-consistently solved with $\frac{\partial m_\alpha}{\partial t} = 0$ by numerical or graphical methods in order to obtain the stationary activity, because $\mu_\alpha(m_E, m_I, m_x)$ and $\sigma_\alpha(m_E, m_I, m_x)$ depend on $m_\alpha \forall \alpha \in \{E, I, X\}$ themselves. We here employ the algorithm hybrd and hybrj from the MINPACK package, implemented in scipy (version 0.9.0) [60] as the function scipy.optimize.fsolve.

2.3 Linearized equation for correlations and susceptibility. In general, the term $\langle F_k(\mathbf{n})\, \delta n_l \rangle$ in (3) couples moments of arbitrary order, resulting in a moment hierarchy [53]. Here we only determine an approximate solution. Since the single synaptic amplitudes J_{ki} are small, we linearize the effect of a single synaptic input. We apply the linearization to the two terms of the form $\langle F_k(\mathbf{n})\, \delta n_l \rangle$ on the right hand side of (3). In the recurrent network, the activity of each neuron in the vector $\mathbf{n}$ may be correlated to the activity of any other neuron n_i. Therefore, the input h_k sensed by neuron k not only depends on n_l directly, but also indirectly through the correlations of n_l with any of the other neurons n_i that project to neuron k. We need to take this dependence into account in the linearization. Considering the effect of one particular input n_i explicitly one gets

$$\begin{aligned} \langle F_k(\mathbf{n}) \delta n_l \rangle &= \langle H(h_k - \theta)\, \delta n_l \rangle \\ &= \langle H(h_{k\setminus n_i} + J_{ki} - \theta)\, n_i \delta n_l + H(h_{k\setminus n_i} - \theta)\, (1 - n_i)\, \delta n_l \rangle \\ &= \langle (H(h_{k\setminus n_i} + J_{ki} - \theta) - H(h_{k\setminus n_i} - \theta))\, n_i \delta n_l \rangle + \langle H(h_{k\setminus n_i} - \theta)\, \delta n_l \rangle. \end{aligned}$$

The first term $\langle (H(h_{k\setminus n_i} + J_{ki} - \theta) - H(h_{k\setminus n_i} - \theta))\, n_i \delta n_l \rangle$ already contains two factors n_i and δn_l, so it takes into account second order moments. Performing the expansion for the next input would yield terms corresponding to correlations of higher order, which are neglected here. This amounts to the assumption that the remaining fluctuations in $h_{k\setminus n_i}$ are independent of n_i and n_l, and we again approximate them by a Gaussian random variable $x \sim \mathcal{N}(\mu_k, \sigma_k)$ with mean $\mu_k = \langle h_k \rangle$ and variance $\sigma_k^2 = \langle \delta h_k^2 \rangle$, so $\langle (H(x + J_{ki} - \theta) - H(x - \theta)) \rangle_x \langle n_i \delta n_l \rangle_{\mathbf{n}} \simeq S(\mu_k, \sigma_k)\, J_{ki}\, \langle n_i \delta n_l \rangle_{\mathbf{n}} + O(J_{ki}^2)$. Here we used the smallness of the synaptic weight J_{ki} and replaced the difference by the derivative $S(\mu_k, \sigma_k) = \left. \frac{\partial \langle H(x+J) \rangle_{x \sim \mathcal{N}(\mu_k, \sigma_k)}}{\partial J} \right|_{J=0}$, which has the form of a susceptibility. Using the explicit expression for the Gaussian integral (7), the susceptibility is exactly

$$S(\mu_k, \sigma_k) \quad = \quad \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{(\mu_k - \theta)^2}{2\sigma_k^2}}. \tag{8}$$

The same expansion holds for the remaining inputs to cell k. With $\langle n_i \delta n_l \rangle = \begin{cases} a_i & \text{for } i = l \\ c_{il} & \text{for } i \neq l \end{cases}$, the equation for the pairwise correlations (3) in linear approximation takes the form

$$2c_{kl} \quad = \quad S(\mu_k, \sigma_k) \left(\sum_j J_{kj} c_{jl} + J_{kl} a_l \right) + S(\mu_l, \sigma_l) \left(\sum_j J_{lj} c_{jk} + J_{lk} a_k \right), \tag{9}$$

corresponding to eq. (6.8) in [33] and eqs. (31)-(33) in [22, Supplementary material]. Note, however, that the linearization used in [33] relies on the smoothness of the gain function due to additional local noise, whereas here and in [22, Supplementary material] a Heaviside gain function is used and only the existence of noise generated by the network itself justifies the linearization. If the input to each neuron is homogeneous, i.e. $\mu_k = \mu_\alpha$ and $\sigma_k = \sigma_\alpha$ for all neurons k in population α, a structurally similar equation connects the correlations $c_{\alpha\beta} = \frac{1}{N_\alpha N_\beta} \sum_{k \in \alpha, l \in \beta, k \neq l} c_{kl}$ averaged over disjoint pairs of neurons belonging to two (possibly identical) populations α,β with the population averaged variances $a_\alpha = \frac{1}{N_\alpha} \sum_{k \in \alpha} a_k$

$$\begin{aligned} 2c_{\alpha\beta} \quad &= \quad \sum_{\gamma \in \{E,I,X\}} (w_{\alpha\gamma} c_{\gamma\beta} + w_{\beta\gamma} c_{\gamma\alpha}) + w_{\alpha\beta} \frac{a_\beta}{N_\beta} + w_{\beta\alpha} \frac{a_\alpha}{N_\alpha} \\ \text{with } w_{\alpha\beta} \quad &= \quad S(\mu_\alpha, \sigma_\alpha)\, J_{\alpha\beta}\, K_{\alpha\beta}. \end{aligned} \tag{10}$$

In deriving the last expression, we replaced variances of individual neurons and correlations between individual pairs by their respective population averages and counted the number of connections. This equation corresponds to eqs. (9.14)-(9.16) in [33] (which lack, however, the external population X, and note the typo in the first term in line 2 of 9.16, which should read $-\frac{1}{2} \bar{J}_{EI} C_{II}(0)$) and eqs. (36) in [22, Supplementary material]. Written in matrix form (10) reads

$$\begin{aligned} A \begin{pmatrix} c_{EE} \\ c_{EI} \\ c_{II} \end{pmatrix} \quad &= \quad B \begin{pmatrix} \frac{a_E}{N_E} \\ \frac{a_I}{N_I} \end{pmatrix} + C \begin{pmatrix} c_{EX} \\ c_{IX} \end{pmatrix} \\ D \begin{pmatrix} c_{EX} \\ c_{IX} \end{pmatrix} \quad &= \quad E \frac{a_X}{N_X}, \end{aligned} \tag{11}$$

resulting in (19) of the main text, where we defined

$$
\begin{aligned}
A &= \begin{pmatrix} 2-2w_{EE} & -2w_{EI} & 0 \\ -w_{IE} & 2-(w_{EE}+w_{II}) & -w_{EI} \\ 0 & -2w_{IE} & 2-2w_{II} \end{pmatrix} \\
B &= \begin{pmatrix} 2w_{EE} & 0 \\ w_{IE} & w_{EI} \\ 0 & 2w_{II} \end{pmatrix} \qquad C = \begin{pmatrix} 2w_{EX} & 0 \\ w_{IX} & w_{EX} \\ 0 & 2w_{IX} \end{pmatrix} \\
D &= \begin{pmatrix} 2-w_{EE} & -w_{EI} \\ -w_{IE} & 2-w_{II} \end{pmatrix} \qquad E = \begin{pmatrix} w_{EX} \\ w_{IX} \end{pmatrix}.
\end{aligned} \tag{12}
$$

The explicit solution of the latter system of (11) is

$$
\begin{pmatrix} c_{XE} \\ c_{XI} \end{pmatrix} = \frac{1}{(2-w_{EE})(2-w_{II})-w_{EI}w_{IE}} \begin{pmatrix} (2-w_{II})w_{EX}+w_{EI}w_{IX} \\ (2-w_{EE})w_{IX}+w_{IE}w_{EX} \end{pmatrix} \frac{a_X}{N_X}. \tag{13}
$$

2.4 Mean-field theory including finite-size correlations. The mean-field solution presented in **"Mean-field solution"** assumes that correlations among the neurons in the network are negligible. This assumption enters the expression (6) for the variance of the input to a neuron. Having determined the actual magnitude of the correlations in (11), we are now able to state a more accurate approximation in which we take these correlations into account, modifying the expression for the variance of the field h_α

$$
\begin{aligned}
\sigma_\alpha^2 &= \sum_{\beta\in\{E,I,X\}} K_{\alpha\beta}J_{\alpha\beta}^2 m_\beta(1-m_\beta) + \sum_{\beta,\gamma\in\{E,I,X\}} (KJ)_{\alpha\beta}(KJ)_{\alpha\gamma}c_{\beta\gamma} \\
\text{with } (KJ)_{\alpha\beta} &\equiv K_{\alpha\beta}J_{\alpha\beta}.
\end{aligned} \tag{14}
$$

This correction suggests an iterative scheme: Initially we solve the mean-field equation (7) assuming $c_{\alpha\beta} = 0$ (hence σ_α given by (6)). In each step of the iteration we then calculate the correlations by (11), compute the mean-field solution of (7) and the susceptibility $S(\mu_\alpha, \sigma_\alpha)$ (8), taking into account the correlations (14) determined in the previous step. These steps are iterated until the solution $(m_\alpha, c_{\alpha\beta} \quad \forall\alpha,\beta)$ converges. We use this approach to determine the correlation structure in Figure 3, where we iterated until the solution became invariant up to a residual absolute difference of 10^{-15}. A comparison of the distribution of the total synaptic input h_E at the end of the iteration with a Gaussian distribution with parameters μ_E and σ_E is shown in Figure 3F.

3 Results

Our aim is to investigate the role of recurrence and external input on the magnitude and structure of cross correlations between the activities in a recurrent random network. Denoting the activity of neuron i as $n_i(t)$ and its mean activity as $m_i = \langle n_i(t)\rangle_t$, the (zero time lag) covariance of the activities of a pair (i,j) of neurons is defined as $c_{ij} = \langle \delta n_i(t)\delta n_j(t)\rangle_t$, where $\delta n_i(t) = n_i(t) - m_i$ is the deviation of neuron i's activity from expectation and the average $\langle\rangle_t$ is over time and realizations of the stochastic activity. We employ the established recurrent neuronal network model of binary neurons in the balanced regime [50]. The binary dynamics has the advantage to be more easily amendable to analytical treatment than spiking dynamics. A method to calculate the pairwise correlations exists since long [33]. The choice of binary dynamics moreover renders our results directly comparable to the recent findings on decorrelation in such networks [22]. Our model consists of three populations of neurons, one excitatory and one inhibitory population which together represent the local network, and an external population providing additional excitatory drive to the local network, as illustrated in Figure 2. The external population may either be conceived as representing input into the local circuit from remote areas or as representing sensory input. The external population contains N_X neurons, which are pairwise uncorrelated and have a stochastic activity with mean m_X. Each neuron in population $\alpha \in \{E, I\}$ within the local network draws

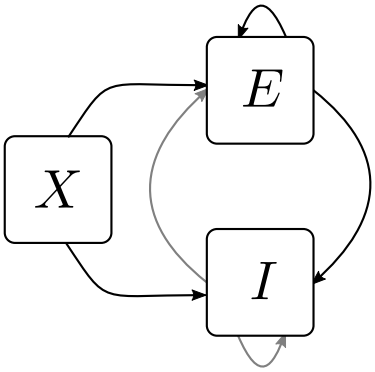

Figure 2: Recurrent local network of two populations of excitatory (E) and inhibitory (I) neurons driven by a common external population (X). The external population X delivers stochastic activity to the local network. The local network is a recurrent Erdős-Rényi random network with homogeneous synaptic amplitudes of weight $J_{\alpha\beta}$ coupling a neuron in population β to a neuron in population α, for $\alpha, \beta \in \{E, I\}$ and same parameters for all neurons. There are $N = 8192$ neurons in both the excitatory and the inhibitory population. The connection probability is $p = 0.2$, and each neuron in population α receives the same number $K = pN$ of excitatory and inhibitory incoming synapses. The size N_X of the external population determines the amount of shared input received by each pair of cells in the local network. The neurons are modeled as binary units with a hard threshold θ.

$K = pN$ connections randomly from the finite pool of N_X external neurons. N_X therefore determines the number of shared afferents received by each pair of cells from the external population with on average K^2/N_X common synapses. In the extreme cases $N_X = K$ all neurons receive exactly the same input, for large $N_X \to \infty$ the fraction of shared external input approaches 0. The common fluctuating input received from the finite-sized external population hence provides a signal imposing pairwise correlations, the amount of which is controlled by the parameter N_X.

3.1 Correlations are driven by intrinsic and external fluctuations. To explain the correlation structure observed in a network with external inputs (Figure 2), we here extend the existing theory of pairwise correlations [33] to include the effect of externally imposed correlations. The global behavior of the network can be studied with help of the mean-field equation (7) for the population-averaged mean activity $m_\alpha = N_\alpha^{-1} \sum_i \langle n_i^\alpha \rangle$

$$m_\alpha = \frac{1}{2}\mathrm{erfc}\left(\frac{\theta - \mu_\alpha}{\sqrt{2}\sigma_\alpha}\right) = \Phi(\mu_\alpha, \sigma_\alpha), \tag{15}$$

where the fluctuations of the input h_α to a neuron in population α are to good approximation Gaussian with the moments

$$\begin{aligned} \mu_\alpha = \langle h_\alpha \rangle &= \sum_\beta K_{\alpha\beta} J_{\alpha\beta} m_\beta \\ \sigma_\alpha^2 = \langle \delta h_\alpha^2 \rangle &= \sum_\beta K_{\alpha\beta} J_{\alpha\beta}^2 m_\beta (1 - m_\beta). \end{aligned} \tag{16}$$

In order to determine the average activities in the network, the mean-field equation (15) needs to be solved self-consistently, as the right hand side depends on the mean activities m_α through (16), as explained in **"Mean-field theory including finite-size correlations"**. Here $K_{\alpha\beta}$ denotes the number of connections from population β to α, and $J_{\alpha\beta}$ their average synaptic amplitude. Once the mean activity in the network has been determined, we can ask for the structure of correlations. For simplicity we focus on the zero time lag correlation, $c_{ij} = \langle \delta n_i(t) \delta n_j(t) \rangle_t$, where $\delta n_i(t) = n_i(t) - \langle n_i \rangle_t$ is the deflection of neuron i's activity from baseline and $a_i = \langle \delta n_i^2(t) \rangle_t = \langle n_i \rangle_t (1 - \langle n_i \rangle_t)$ is the variance of neuron i's activity. Starting from the master equation for the network of binary neurons, in **"Methods"** for completeness

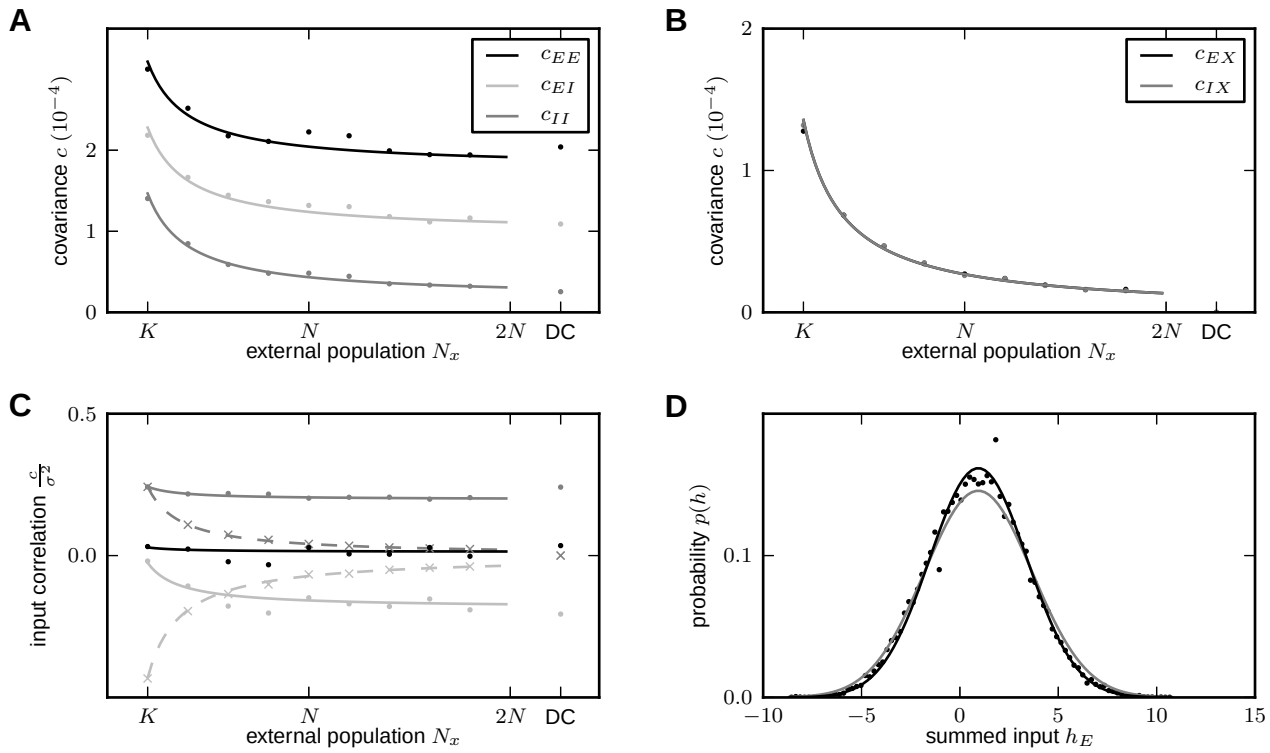

Figure 3: Correlations in a network of three populations as illustrated in Figure 2 in dependence of the size N_x of the external population. Each neuron in population $\alpha \in \{E, I\}$ receives pN randomly drawn excitatory inputs with weight $J_{\alpha E} = \frac{5}{\sqrt{N}}$, pN randomly drawn inhibitory inputs of weight $J_{\alpha I} = -\frac{10}{\sqrt{N}}$ and pN external inputs of weight $J_{\alpha X} = \frac{5}{\sqrt{N}}$ (homogeneous random network with fixed in-degree). **A** Correlations averaged over pairs of neurons within the local network (17). Dots are results of direct simulation over $T = 30\,\mathrm{s}$ averaged over $(N/2)^2$ pairs of neurons and curves show the analytical result (19). The point "DC" shows the correlation structure emerging if the drive from the external population is replaced by a constant value with the same mean. **B** Correlations between neurons within the local network and the external population averaged over pairs of neurons (same labeling as in A). **C** Correlation between the inputs to a pair of cells in the network decomposed into the contributions due to shared inputs c_{shared} (gray, eq. 20) and due to correlations c_{corr} (light gray, eq. 21). Dashed curves and St. Andrew's Crosses show the contribution due to external inputs, solid curves and dots show the contribution from local inputs. The sum of all components is shown by black dots and curve. Curves are theoretical results based on (19), (20), and (21), symbols are obtained from simulation. **D** Probability distribution of the fluctuating input h_E to a single neuron in the excitatory population. Dots show the histogram obtained from simulation binned over the interval $[\min(h_E), \max(h_E)]$ with a bin size of $-2J_{\alpha I}$. The gray curve is the prediction of a Gaussian distribution obtained from mean-field theory neglecting correlations, with mean and variance given by (4) and (6), respectively. The black curve takes into account correlations in the afferent signals and has a variance given by (14). Other parameters: simulation resolution $\Delta t = 0.1\,\mathrm{ms}$, synaptic delay $d = \Delta t$, activity measurement in intervals of 1 ms. Threshold of the neurons $\theta = 1$, time constant of inter-update-intervals $\tau = 10\,\mathrm{ms}$. The average activity in the network is $m_E \simeq m_I \simeq m_X = 0.5$.

and consistency in notation we re-derive the self-consistent equation that connects the cross covariances $c_{\alpha\beta}$ averaged over pairs of neurons from population α and β and the variances a_α averaged over neurons from population α

$$\begin{aligned} c_{\alpha\beta} &= \frac{1}{N_\alpha N_\beta} \sum_{k\in\alpha, l\in\beta, k\neq l} c_{kl} \\ a_\alpha &= \frac{1}{N_\alpha} \sum_{k\in\alpha} a_k. \end{aligned} \tag{17}$$

The obtained inhomogeneous system of linear equations (19) reads [33]

$$2c_{\alpha\beta} = \frac{1}{N_\beta} w_{\alpha\beta} a_\beta + \sum_{\gamma\in\{E,I,x\}} w_{\alpha\gamma} c_{\gamma\beta} + \text{transpose}(\alpha \leftrightarrow \beta). \tag{18}$$

Here $w_{\alpha\beta}$ measures the effective linearized coupling strength from population β to population α that depends on the number of connections $K_{\alpha\beta}$ from population β to α, their average synaptic amplitude $J_{\alpha\beta}$ and the susceptibility S_α of neurons in population α as $w_{\alpha\beta} = S(\mu_\alpha, \sigma_\alpha)\, K_{\alpha\beta}\, J_{\alpha\beta}$. The susceptibility $S(\mu_\alpha, \sigma_\alpha)$ given by (8) quantifies the influence a fluctuation in the input to a neuron in population α has on the output. It depends on the working point $(\mu_\alpha, \sigma_\alpha)$ of the neurons in population α. The autocorrelations a_E, a_I and a_X are the inhomogeneity in the system of equations, so they drive the correlations, as pointed out earlier [33]. This is in line with the linear theories [61, 28] for leaky integrate-and-fire model neurons, where cross-correlations are proportional to the auto-correlations. The system of equations (18) is identical to [33, eqs. (9.14)-(9.16)]. Note that this description holds for finite-sized networks. With the symmetry $c_{EI} = c_{IE}$, (18) can be written in matrix form as

$$\begin{aligned} A \begin{pmatrix} c_{EE} \\ c_{EI} \\ c_{II} \end{pmatrix} &= B \begin{pmatrix} \frac{a_E}{N_E} \\ \frac{a_I}{N_I} \end{pmatrix} + C \begin{pmatrix} c_{EX} \\ c_{IX} \end{pmatrix} \\ D \begin{pmatrix} c_{EX} \\ c_{IX} \end{pmatrix} &= E \frac{a_X}{N_X}, \end{aligned} \tag{19}$$

with the explicit forms of the matrices $A, \ldots, E$ given in (12). This system of linear equations can be solved by elementary methods. From the structure of the equations it follows, that the correlations between the external input and the activity in the network, c_{EX} and c_{IX}, are independent of the other correlations in the network. They are solely determined by the solution of the system of equations in the second line of (19), driven by the fluctuations of the external drive a_X/N_X. The correlations among the neurons within the network are given by the solution of the first system in (19). They are hence driven by two terms, the fluctuations of the neurons within the network proportional to a_E/N_E and a_I/N_I and the correlations between the external population and the neurons in the network, c_{EX} and c_{IX}.

The second line of (19) shows that all correlations depend on the size N_X of the external population. Since the number $K = pN$ of randomly drawn afferents per neuron from this population is constant, the mean number of shared inputs to a pair of neurons is K^2/N_X. In the extreme case $N_X = K$ on the left of Figure 3 all neurons receive exactly identical input. If the recurrent connectivity would be absent, we would hence have perfectly correlated activity within the local network, the covariance between two neurons would be equal to their variance $a_\alpha = m_\alpha(1 - m_\alpha)$, in this particular network $a_\alpha \simeq 0.25$. Figure 3A shows that the covariance in the recurrent network is much smaller; on the order of 10^{-4}. The reason is the recently reported mechanism of decorrelation [22], explained by the negative feedback in inhibition-dominated networks [17]. Increasing the size of the external population decreases the amount of shared input, as seen in Figure 3C. In the limit where the external drive is replaced by a constant value (shown as point "DC"), the external drive does consequently not contribute to correlations in the network. Figure 3A shows that the relative position of the three curves does not change with N_X. The

overall offset, however, changes. This can be understood by inspecting the analytical result (19). The solution of this system of linear equations is a superposition of two contributions. One is due to the externally imposed fluctuations, proportional to a_X/N_X the other is due to fluctuations generated within the local network, proportional to a_X/N_E and a_I/N_I. Varying the size of the external population only changes the external contribution, causing the variation in the offset, while the internal contribution, causing the splitting between the three curves, stays constant. In the extreme case (point DC, $a_X = 0$) we still observe a similar structure. The slightly larger splitting is due to the reduced variance σ_α^2 in the single neuron input, which consequently increases the susceptibility S_α (8).

Figure 3D shows the probability distribution of the input h_α to a neuron in population $\alpha = E$. The histogram is well approximated by a Gaussian. The first two moments of this Gaussian are μ_α and σ_α^2 given by (16), if correlations among the afferents are neglected. This approximation is shown to deviate from the result of direct simulation. Taking into account the correlations among the afferents affects the variance in the input according to (14). The latter approximation is a better estimate of the input statistics, as shown in Figure 3D. This improved estimate can be accounted for in the solution of the mean-field equation (15), which in turn via the susceptibility S_α affects the correlations. Iterating this procedure until convergence, as explained in **"Mean-field theory including finite-size correlations"**, yields the analytical results presented in Figure 3.

3.2 Cancellation of input correlations. We would like to understand how the structure of correlations relates to the earlier report of fast tracking [50, 52]. Small correlations observed in balanced recurrent networks were explained by the property of recurrent networks to track their input on a fast time-scale [22, their eq. (2)]. Figure 4A shows the population activities in a network of three populations for fixed numbers of neurons $N_x = N_E = N_I = N$ and symmetric connectivity as in Figure 3. The deflections of the excitatory and the inhibitory population partly resemble those of the external drive to the network, but partly the fluctuations seem to be independent. Our theoretical result for the correlation structure explains this result (19): the fluctuations in the network are not only driven by external input (proportional to a_X), but are also driven by the fluctuations generated within the local populations (proportional to a_E and a_I). The idea of fast tracking of the external signal was derived from the observation that the fluctuations in the population-averaged input $h_\alpha = \frac{1}{N_\alpha}\sum_{i\in\alpha} h_i$ are suppressed [22]. This suppression can be observed by decomposing the input h_α to the population α into contributions from excitatory (including external neurons) and from inhibitory cells, $h_{\alpha E} = (KJ)_{\alpha E} n_E + (KJ)_{\alpha X} n_X$ and $h_{\alpha I} = (KJ)_{\alpha I} n_I$, respectively, where we used the short hand $(KJ)_{\alpha\beta} = K_{\alpha\beta} J_{\alpha\beta}$. As shown in Figure 4E, the contributions of excitation and inhibition cancel each other so that the total input fluctuates close to the threshold (here $\theta = 1$) of the neurons: the network is in the balanced state [50]. Moreover, this cancellation not only holds for the mean value, but also for fast fluctuations, which are consequently reduced in the sum h_α compared to the individual components $h_{\alpha E}$ and $h_{\alpha I}$ (Figure 4E). We will now show that this suppression of fluctuations directly implies a relation for the correlation $\langle \delta h_i \delta h_j \rangle$ between the inputs to a pair (i, j) of individual neurons. There are two distinct contributions to this correlation $\langle \delta h_i \delta h_j \rangle = c_{\text{shared},\alpha} + c_{\text{corr},\alpha}$, one due to common inputs shared by the pair of neurons (both neurons i, j assumed to belong to population α)

$$c_{\text{shared},\alpha} = \sum_{\beta\in\{E,I,X\}} (KJ)_{\alpha\beta}^2 \frac{a_\beta}{N_\beta} \tag{20}$$

and one due to the correlations between afferents

$$c_{\text{corr},\alpha} = \sum_{\beta,\gamma\in\{E,I,X\}} (KJ)_{\alpha\beta}(KJ)_{\alpha\gamma} c_{\beta\gamma}. \tag{21}$$

Figure 4C shows these two contributions to be of opposite sign but approximately of same magnitude. Figure 3C shows a further decomposition of the input correlation into contributions due to the external sources and due to connections from within the local network. The sum of all components is much smaller

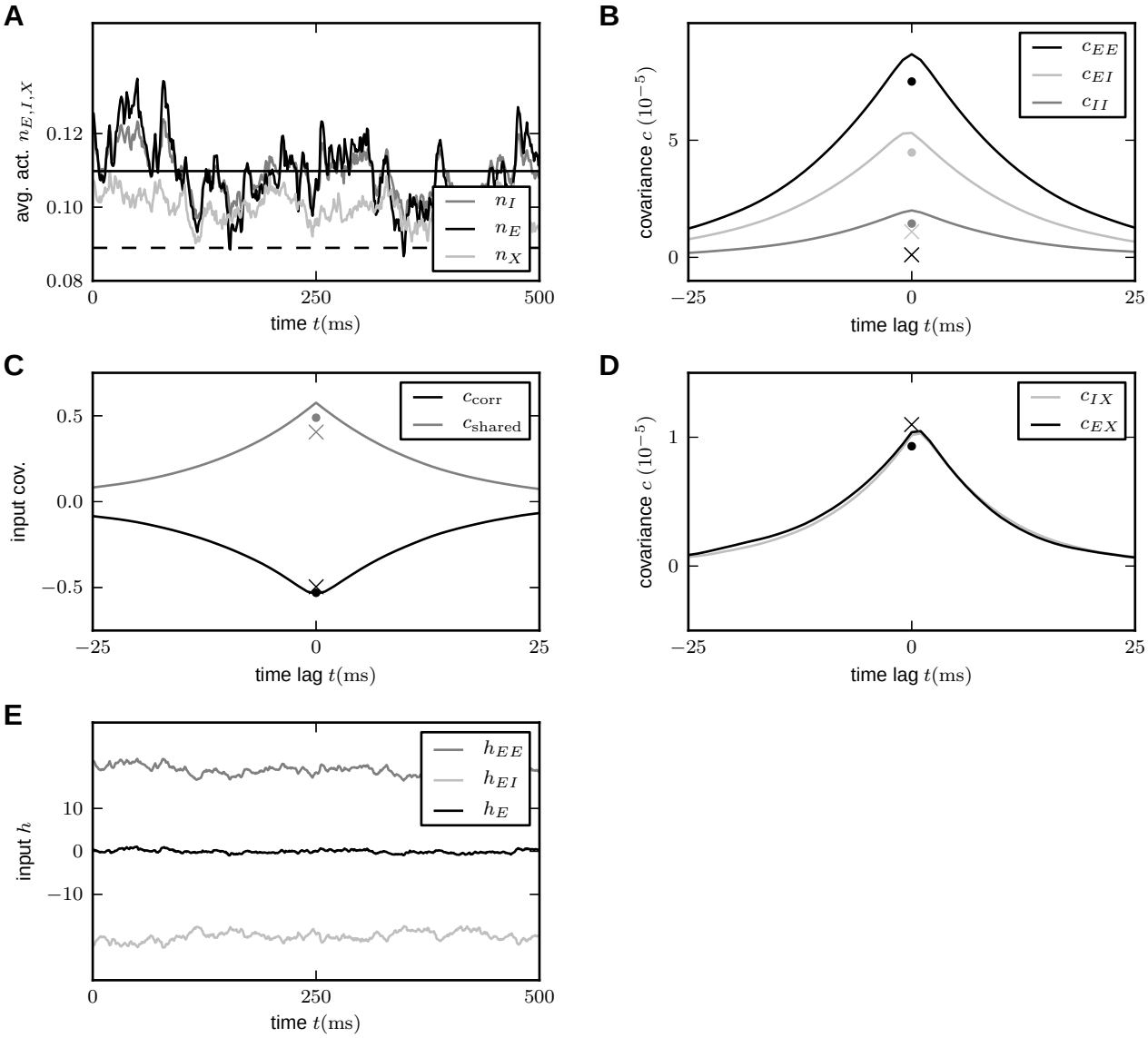

Figure 4: Activity in a network of $3N = 3\times 8192$ binary neurons with synaptic amplitudes $J_{\alpha E} = J_{\alpha X} = 5/\sqrt{N}$, $J_{\alpha I} = -10/\sqrt{N}$ depending exclusively on the type of the sending neuron (E or I). Each neuron receives $K = pN$ randomly drawn inputs (fixed in-degree). **A** Population averaged activity (black E, gray I, light gray X). Analytical prediction (5) for the mean activities $m_E = m_I$ (dashed horizontal line) and numerical solution of mean field equation (7) (solid horizontal line). **B** Cross covariance between excitatory neurons (black), between inhibitory neurons (gray), and between excitatory and inhibitory neurons (light gray). Theoretical results (19) shown as dots. St. Andrew's Crosses indicate the theoretical prediction of [22, suppl. eqs. 38,39]. **C** Correlation between the input currents to a pair of excitatory neurons. The black curve is the contribution due to pairwise correlations c_{corr}, the gray curve is the contribution of shared input c_{shared}. The symbols show the theoretical expectation (20) and (21) based on [22] (crosses) and based on (19) (dots). **D** Similar as B, but showing the correlations between external neurons and neurons in the excitatory and inhibitory population. Note that both theories yield $c_{EX} = c_{IX}$, so for each theory ([22] crosses, (19) dots) only the symbol for c_{EX} is visible. **E** Contributions h_{EE} (gray) due to excitatory synapses and h_{EI} (light gray) due to inhibitory synapses to the input h_E averaged over all excitatory neurons. Duration of simulation $T = 100\,\mathrm{s}$, mean activity $m_X = 0.1$, $m_E \simeq m_I \simeq 0.11$, other parameters as in Figure 3.

than each individual component. This cancellation can be understood from the observation of the small fluctuations in the population-averaged input $\langle \delta h_\alpha^2 \rangle \simeq 0$, because

$$
\begin{aligned}
0 \simeq \langle \delta h_\alpha^2 \rangle = \left\langle \left(\sum_{\beta \in \{E,I,X\}} (KJ)_{\alpha\beta} \delta n_\beta \right)^2 \right\rangle &= \sum_{\beta,\gamma \in \{E,I,X\}} (KJ)_{\alpha\beta} (KJ)_{\alpha\gamma} \langle \delta n_\beta \delta n_\gamma \rangle \\
&= \sum_{\beta \in \{E,I,X\}} (KJ)_{\alpha\beta}^2 \frac{a_\beta}{N_\beta} + \sum_{\beta,\gamma \in \{E,I,X\}} (KJ)_{\alpha\beta} (KJ)_{\alpha\gamma} \, c_{\beta\gamma} \\
&= c_{\mathrm{shared},\alpha} + c_{\mathrm{corr},\alpha},
\end{aligned}
\tag{22}
$$

where in the second step we used the general relation between the covariance $\langle \delta n_\beta \delta n_\gamma \rangle$ among two population averaged signals n_β and n_γ and the population-averaged variance a_β and pairwise averaged covariances $c_{\beta\gamma}$, which reads [17, cf. eq. (1)]

$$
\begin{aligned}
\langle \delta n_\beta \delta n_\gamma \rangle &= \left\langle \frac{1}{N_\beta N_\gamma} \sum_{i \in \beta, j \in \gamma} \delta n_i \delta n_j \right\rangle \\
&= \delta_{\beta\gamma} \frac{1}{N_\beta^2} \sum_{i \in \beta} \langle \delta n_i^2 \rangle + \frac{1}{N_\beta N_\gamma} \sum_{i \in \beta, j \in \gamma, i \neq j} \langle \delta n_i \delta n_j \rangle \\
&= \delta_{\beta\gamma} \frac{1}{N_\beta} a_\beta + c_{\beta\gamma}.
\end{aligned}
\tag{23}
$$

This suppression of fluctuations in the population-averaged input is a consequence of the overall negative feedback in these networks: a fluctuation δh_α of the population averaged input h_α causes a response in network activity which is coupled back with a negative sign, counteracting its own cause and hence suppressing the fluctuation δh_α. Expression (22) is an algebraic identity showing that hence also correlations between the total inputs to a pair of cells must be suppressed. This argument also shows why the suppression of input-correlations does not rely on a balance between excitation and inhibition; it can as well be observed in purely inhibitory networks [17, cf. text following eq. (21) therein], where the overall negative feedback suppresses population fluctuations δh_α in exactly the same manner.

It is easy to see that the cancellation condition (22) does not uniquely determine the correlations in the network. Figure 5 shows the same measures of activity as Figure 4, but for the network connectivity used in [22, their Fig. 2]. Except for J_{II} and J_{IX} the parameters are the same as in Figure 4. Moreover, we distributed the number of incoming connections K per neuron according to a binomial distribution, as in the original publication. The cancellation of fluctuations on the input side is evident from Figure 5C and E. Comparing panels B of Figure 4 and Figure 5 we observe that this cancellation, however, does not result in the same correlation structure: In Figure 4 the correlations fulfill $c_{EE} > c_{EI} > c_{II}$, while in Figure 5 the relation is $c_{EI} \simeq c_{EE} > c_{II}$. This also explains why the attempt to derive the structure of correlations from $\langle \delta h^2 \rangle \simeq 0$ fails, as seen in Figure 5B: the predicted correlations [22, Supplementary material, eqs. 38-39] do not coincide with the results obtained by direct simulation. Note, however, that the wrong correlation structure, due to the above discussed relation (22), still fulfills the cancellation condition for correlations on the input side, as shown in Figure 5C. The reason for the failure is the assumption that fluctuations in the network are exclusively driven by the external input. This assumption [22, eq. (2)] is reflected in the expression $c_{EI} \propto a_X$, which is directly proportional to the external fluctuations. What is missing is the contribution due to the intrinsic fluctuations of the local populations (E,I), as they appear in (19). This can be shown explicitly by setting $a_E = 0$ and $a_I = 0$ in (19), resulting in a similar prediction for c_{EI}, as shown in Figure 5B (plus symbol). The remaining deviation between both theories is due to the different susceptibilities S used by the two approaches. Note that the full theory (19) predicts the structure of correlations with only minor deviations. In summary, the cancellation condition imposes a constraint on the structure of correlations but is not sufficient as a unique determinant.

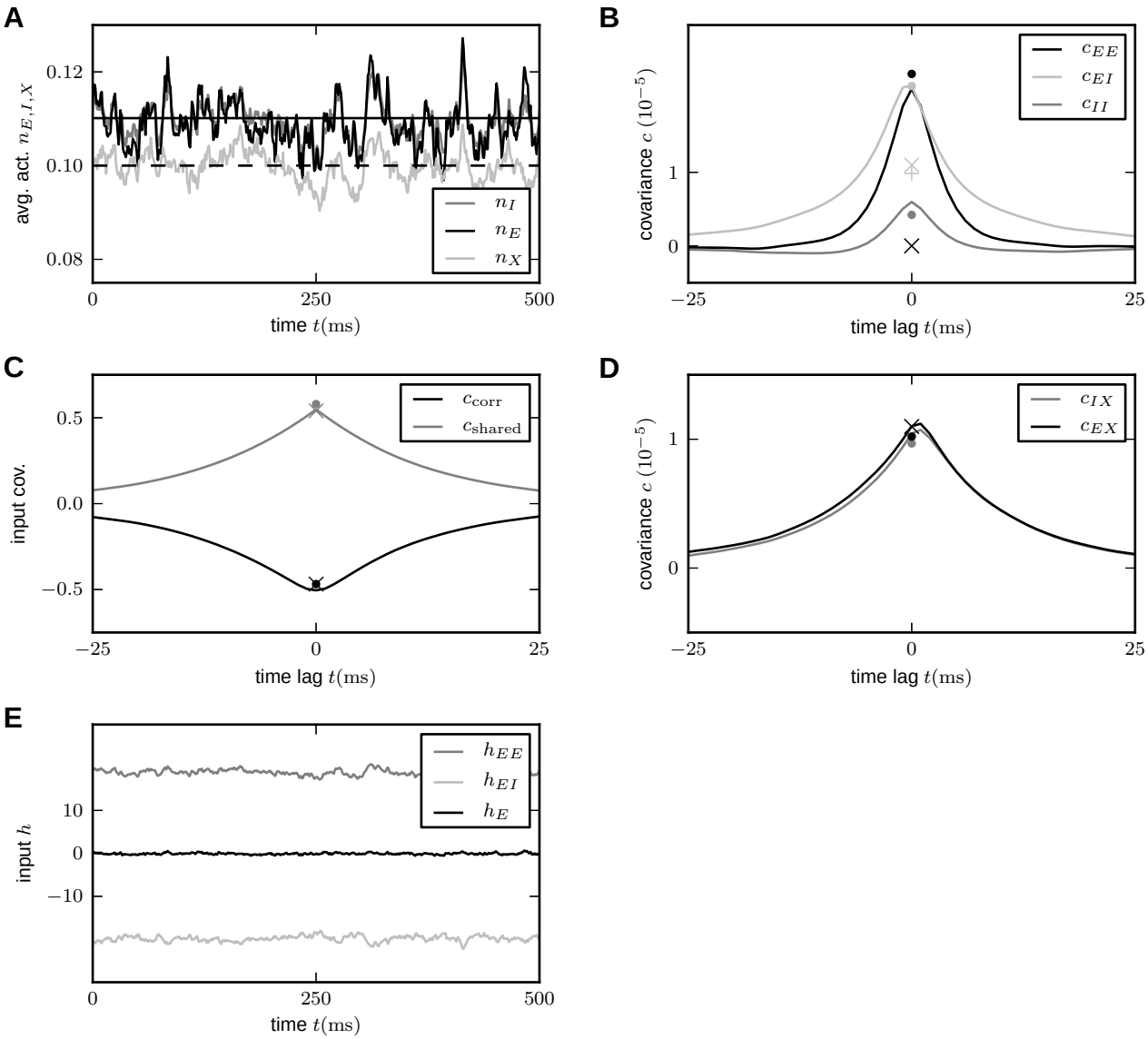

Figure 5: Activity in a network of $3N = 3\times 8192$ binary neurons as described in [22], with $J_{EE} = 5/\sqrt{N}$, $J_{EI} = -10/\sqrt{N}$, $J_{IE} = 5/\sqrt{N}$, $J_{II} = -9/\sqrt{N}$, $J_{EX} = 5/\sqrt{N}$, $J_{IX} = 4/\sqrt{N}$. Number K of synaptic inputs binomially distributed as $K \sim B(N, p)$, with connection probability $p = 0.2$. **A** Population averaged activity (black E, gray I, light gray X). Analytical prediction (5) for the mean activities $m_E = m_I$ (dashed horizontal line) and numerical solution of mean field equation (7) (solid horizontal line). **B** Cross correlation between excitatory neurons (black curve), between inhibitory neurons (gray curve), and between excitatory and inhibitory neurons (light gray curve) obtained from simulation. St. Andrew's Crosses show the theoretical prediction from [22, suppl. eqs. 38,39] (prediction yields $c_{EE} = c_{II} = 0$, so only one cross shown). Dots show the theoretical prediction (19). The plus symbol shows the prediction for the correlation c_{EI} if the terms proportional to a_E and a_I are set to zero. **C** Correlation between the input currents to a pair of excitatory neurons. Contribution due to pairwise correlations $c_{\mathrm{corr},E}$ (black curve) and due to shared input $c_{\mathrm{shared},E}$ (gray curve). Symbols show the theoretical predictions based on [22] (crosses) and based on (19) (dots). **D** Similar as B, but showing the correlations between external neurons and neurons in the excitatory and inhibitory population. **E** Fluctuating input h_E averaged over the excitatory population (black), separated into contributions from excitatory synapses h_{EE} (gray) and from inhibitory synapses h_{EI} (light gray). Duration of simulation $T = 100\,\mathrm{s}$, mean activity $m_X = 0.1$, $m_E \simeq m_I \simeq 0.11$, other parameters as in Figure 3.

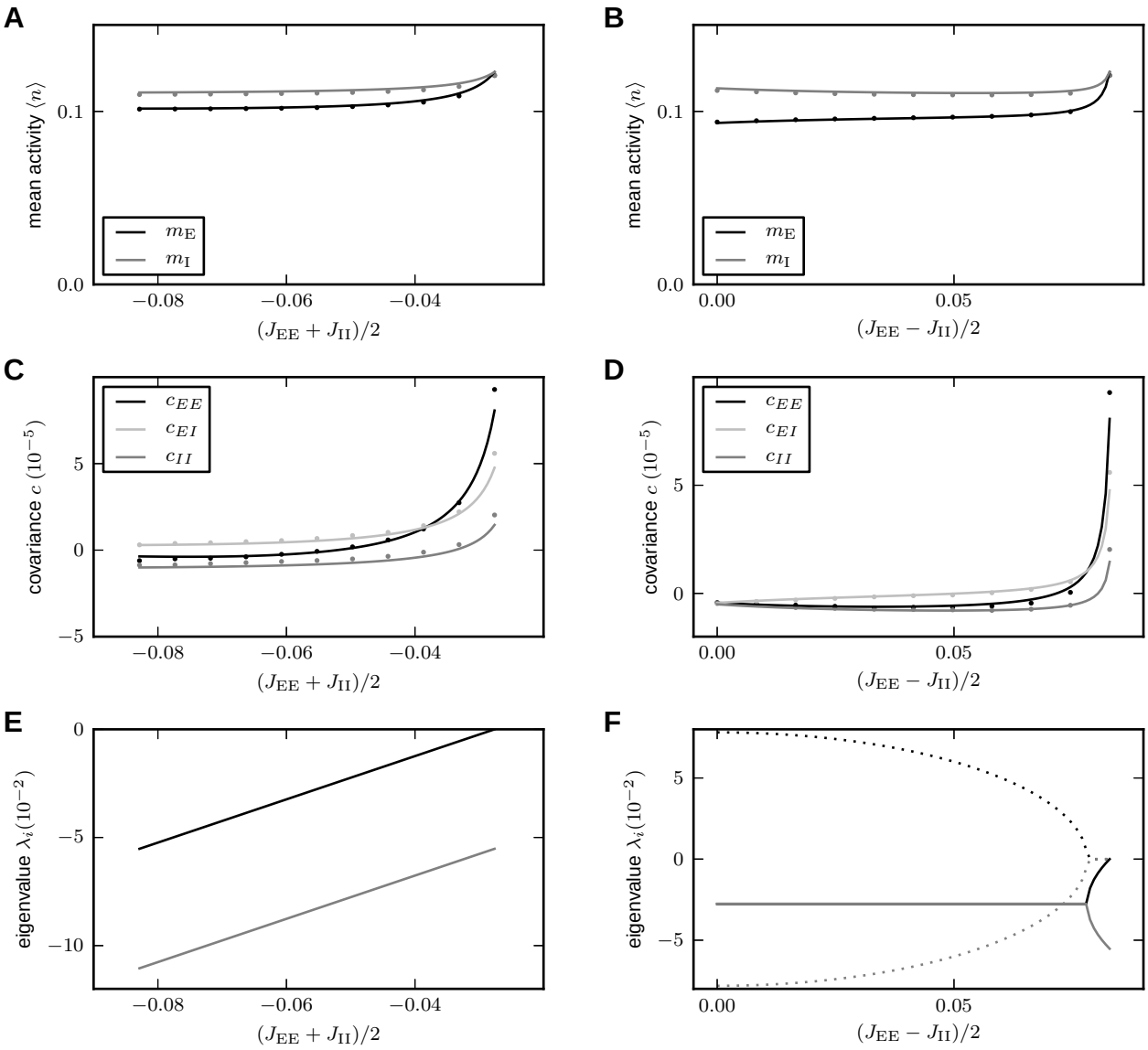

Figure 6: Connectivity structure determines correlation structure. In the left column (A,C,E) we vary $c_1 = (J_{EE} + J_{II})/2$, in the right column (B,D,F) $c_2 = (J_{EE} - J_{II})/2$. **A**,**B** Mean activity in the network as a function of the structural parameters c_1 and c_2, respectively. **C**,**D** Correlations averaged over pairs of neurons. Dots obtained from direct simulation, solid curves given by theory (19) **E**,**F** Eigenvalues (25) of the population-averaged connectivity matrix; solid curves show the real part, dashed curves the imaginary part.

3.3 Influence of connectivity on the correlation structure. Comparing Figure 4B and Figure 5B, the structure of correlations is obviously different. In Figure 4B, the structure is $c_{EE} > c_{EI} > c_{II}$, whereas in Figure 5B the relation is $c_{EI} \simeq c_{EE} > c_{II}$. The only difference between these two networks are the coupling strengths J_{II} and J_{IX}. In the following we derive a more complete picture of the determinants of the correlation structure. In order to identify the parameters that influence the fluctuations in these networks, it is instructive to study the mean-field equation for the population-averaged activities. Linearizing (15) for small deviations $\delta n_\alpha = n_\alpha - m_\alpha$ of the population-averaged activity n_α from the fixed point m_α, for large networks with $N > K_{\alpha\beta} \gg 1$ the dominant term is proportional to the change of the mean $\delta\mu_\alpha = \sum_\beta (JK)_{\alpha\beta}\delta n_\beta$, because the standard deviation $\delta\sigma_\alpha$ is only proportional to $\sqrt{K_{\alpha\beta}}$. To linear order we hence have a coupled set of two differential equations

$$\begin{aligned} \tau\frac{\partial}{\partial t}\begin{pmatrix}\delta n_E\\ \delta n_I\end{pmatrix} + \begin{pmatrix}\delta n_E\\ \delta n_I\end{pmatrix} &= \begin{pmatrix}w_{EE} & w_{EI}\\ w_{IE} & w_{II}\end{pmatrix}\begin{pmatrix}\delta n_E\\ \delta n_I\end{pmatrix} + \text{noise} + \text{external drive} \qquad (24)\\ \text{with } w_{\alpha\beta} &= S(\mu_\alpha,\sigma_\alpha)\,(KJ)_{\alpha\beta}\\ \text{and } S(\mu_\alpha,\sigma_\alpha) &= \frac{\partial\Phi(\mu_\alpha,\sigma_\alpha)}{\partial\mu_\alpha}. \end{aligned}$$

The dynamics of this coupled set of linear differential equations is determined by the two eigenvalues of the effective connectivity

$$\begin{aligned} \lambda_{1,2} &= \text{eig}\left(\{w_{\alpha\beta}\}\right)\\ &= \frac{w_{EE}+w_{II}}{2} \pm \sqrt{\left(\frac{w_{EE}-w_{II}}{2}\right)^2 + w_{EI}w_{IE}}. \qquad (25) \end{aligned}$$

Due to the presence of the leak term on the left hand side of (24), the fixed point rate is stable only if the real parts of the eigenvalues $\lambda_{1,2}$ are both smaller than 1. In the network with identical input statistics for all neurons the fluctuating input is characterized by the same mean and variance (μ, σ^2) for each neuron. For homogeneous neuron parameters the susceptibility $S_\alpha = S$ is hence the same for both populations $\alpha \in \{E, I\}$. If further the number of synaptic afferents is the same $K_{\alpha\beta} = K$ for all populations, the eigenvalues can be expressed by those of the original connectivity matrix as

$$\begin{aligned} \frac{\lambda_{1,2}}{S(\mu,\sigma)K} &= \frac{J_{EE}+J_{II}}{2} \pm \sqrt{\left(\frac{J_{EE}-J_{II}}{2}\right)^2 + J_{EI}J_{IE}}\\ &= c_1 \pm \sqrt{c_2^2 + J_{EI}J_{IE}}, \end{aligned}$$

where we defined the two parameters c_1 and c_2 which control the location of the eigenvalues. In the left column of Figure 6 we keep J_{EI}, J_{IE}, and c_2 constant and vary $c_1 \in [-c_2, -\sqrt{c_2^2 + J_{EI}J_{IE}}]$, where we choose the maximum value by the condition $\lambda_1 < 0$ and the minimum value by the condition that $J_{EE} \geq 0$ and $J_{II} \leq 0$, leading to $c_1 + c_2 \geq 0$ and $c_1 - c_2 \leq 0$, both fulfilled if $-c_2 \leq c_1 \leq c_2$. Varying c_2 in the right column of Figure 6, the bounds are given by the same condition that $J_{EE} \geq 0$ and $J_{II} \leq 0$, so $c_2 \geq 0$, and the condition for the larger eigenvalue to stay below or equal 0, so $c_2 \in [0, \sqrt{c_1^2 - J_{EI}J_{IE}}]$. In order for the network to maintain similar mean activity, we choose the threshold of the neurons such that the cancellation condition $0 = \sum_{\beta\in\{E,I,X\}}(KJ)_{\alpha\beta}m_\beta - \theta$ is fulfilled for $m_\beta = 0.1$. The resulting average activity is close to this desired value of 0.1 and agrees well to the analytical prediction (15), as shown in Figure 6A, B.

The right-most point in both columns of Figure 6 where one eigenvalue vanishes $\lambda_1 = 0$, results in the same connectivity structure. This is the case for the connectivity with the symmetry $J_{EE} = J_{IE} = J$ and $J_{II} = J_{EI} = -gJ$ (cf. Figure 4), because in this case the population averaged connectivity matrix has two linearly dependent rows, hence a vanishing determinant and thus an eigenvalue 0. As observed in Figure 6C,D at this point the absolute magnitude of correlations is largest. This is intuitively clear as the network has a degree of freedom in the direction of the eigenvector $v_1 = (g, 1)^T$ belonging to

the vanishing eigenvalue $\lambda_1 = 0$. In this direction the system effectively does not feel any negative feedback, so the evolution is as if the connectivity would be absent. Fluctuations in this direction hence become large and are only damped by the exponential relaxation of the neuronal dynamics, given by the left hand side of (24). The time constant of these fluctuations is then solely determined by the time constant of the single neurons, as seen in Figure 4B. From the coefficients of the eigenvector we can further conclude that the fluctuations of the excitatory population are stronger by a factor g than those of the inhibitory population, explaining why $c_{EE} > c_{II}$, and that both populations fluctuate in-phase, so $c_{EI} > 0$, (Figure 6C,D, right most point). Moving away from this point, Figure 6C,D both show that the magnitude of correlations decreases. Comparing the temporal structures of Figure 4B and Figure 5B shows that also the time scale of fluctuations increases. The two structural parameters c_1 and c_2 affect the eigenvalues of the connectivity in a distinct manner. Changing c_1 merely shifts the real part of both eigenvalues, but leaves their relative distance constant, as seen in Figure 6E. For smaller values of c_1 the coupling among excitatory neurons becomes weaker, so their correlations are reduced. At the left most point in Figure 6C the coupling within the excitatory population vanishes, $J_{EE} = 0$. Changing the parameter c_2 has a qualitatively different effect on the eigenvalues, as seen in Figure 6F. At $c_2 = \sqrt{|J_{EI}J_{IE}|}$, the two real eigenvalues merge and for smaller c_2 they turn into a conjugate complex pair. At the left-most point $J_{EE} - J_{II} = 0$, so both couplings within the populations vanish $J_{EE} = J_{II} = 0$. The system then only has coupling from E to I and vice versa. The conjugate complex eigenvalues show that the population activity of the system has oscillatory solutions. This is also called the PING (pyramidal - inhibitory - gamma) mechanism of oscillations in the gamma-range [62]. Figure 6C,D show that for most connectivity structures the correlation structure is $c_{EI} > c_{EE} > c_{II}$, in contrast to our previous finding [17], where we studied the symmetric case (the right-most point), at which the correlation structure is $c_{EE} > c_{EI} > c_{II}$. The comparison of the direct simulation to the theoretical prediction (19) in Figure 6C,D shows that the theory yields an accurate prediction of the correlation structure for all connectivity structures considered here.

4 Discussion

The present work explains the observed pairwise correlations in a homogeneous random network of excitatory and inhibitory binary model neurons driven by an external population of finite size. On the methodological side the work required a minor extension of the seminal theory of correlations in stochastic binary networks by [33]. In particular, we improved the linearization procedure to account for the fluctuations due to the balanced synaptic noise [63]. This enables us to apply the existing theory [33] to neuron models with a hard threshold, without the need to assume the presence of additional noise local to each neuron. We further extend the framework by an iterative procedure taking into account the finitesize fluctuations in the mean-field solution to determine the working point (mean activity) of the network. We find that the iteration converges to predictions for the covariance with higher accuracy than the previous method.

Equipped with these methods we investigate a network driven by correlated input due to shared afferents supplied by an external population. The analytical expressions for the covariances averaged over pairs of neurons show that correlations have two components that linearly superimpose, one caused by intrinsic fluctuations generated within the local network and one caused by fluctuations due to the external population. The size N_X of the external population controls the strength of the correlations in the external input. We find that this external input causes an offset of all pairwise correlations, which decreases with increasing external population size in proportion to the strength of the external correlations ($\propto 1/N_X$). The structure of correlations within the local network, i.e. the differences between correlations for pairs of neurons of different types, is mostly determined by the intrinsically generated fluctuations. These are proportional to the population-averaged variances a_E and a_I of the activity of the neurons in the local network. As a result, the structure of correlations is mostly independent of the external drive, even for the limiting case of an infinitely large external population $N_X \to \infty$ or if the external drive is replaced by a DC signal with the same mean. For the other extreme, when the

size of the external population equals the number of external afferents, $N_X = K$, all neurons receive an exactly identical external signal. We show that the mechanism of decorrelation [22, 17] still holds for these strongly correlated external signals. The resulting correlation within the network is much smaller than expected given the amount of common input. In contrast to an earlier explanation [22], which invokes the network's fast tracking of the external drive [50, 52] as the cause of small correlations, we here show that the cancellation of correlations between the inputs to pairs of neurons is equivalent to a suppression of fluctuations of the population-averaged input. This argument is in line with the earlier explanation that correlations are suppressed by negative feedback on the population level [17]. We further show that the cancellation of correlations does not uniquely determine the structure of correlations; there are different structures of correlations that lead to a cancellation of correlations between the summed inputs. The cancellation of input correlations therefore only constitutes a constraint for the pairwise correlations in the network. This constraint is trivially fulfilled under the assumption of perfect tracking, which assumes input fluctuations δh to vanish completely. Therefore the attempt to obtain the structure of finite-sized correlations from fast tracking fails, although such a theory -by construction- reproduces the cancellation of input correlations [22]. Starting from the fast-tracking assumption, the resulting expressions for the correlations within the network [22, Supplementary, eqs. 38,39] are lacking the locally generated fluctuations as additional sources, shown here to be crucial for shaping the correlation structure. Note, however, that the intermediate result [22, Supplementary, eqs. 31,33] is still correct and identical to [33, eq. 6.8] and to (9).

For a common but special choice of network connectivity where the synaptic weights depend only on the type of the source but not the target neuron, i.e. $J_{EE} = J_{IE}$ and $J_{EI} = J_{II}$ [42], we show that the locally generated fluctuations and correlations are elevated and that the activity only loosely tracks the external input. The correlation structure exhibited is $c_{EE} > c_{EI} > c_{II}$. To systematically investigate the dependence of the correlation structure on the network connectivity, it proves useful to parameterize the structure of the network by two measures differentially controlling the location of the eigenvalues of the connectivity matrix. We find that for a wide parameter regime the correlations change quantitatively, but the correlation structure $c_{EI} > c_{EE} > c_{II}$ remains invariant. The qualitative comparison with experimental observations of [49] hence only constrains the connectivity to be within the one or the other parameter regime.

The presented theory is limited to sufficiently asynchronous and irregular network states. This limitation arises from the linearization procedure, which approximates the summed synaptic input by a Gaussian random variable. The deviations of the theory from direct simulations are stronger at lower mean activity, when the synaptic input fluctuates in the non-linear part of the effective transfer function. The best agreement of theory and simulation is hence obtained for a mean population activity close to $\frac{1}{2}$, where 1 means all neurons are active. For simplicity we consider in the theory networks where neurons have a fixed in-degree. In large homogeneous random networks this is often a good approximation, because the mean number of connections is $pN \propto N$, and its standard deviation $\sqrt{Np(1-p)} \propto \sqrt{N}$ declines relative to the mean. A comparison of this simplifying ansatz to a more realistic simulation with distributed in-degrees in Figure 5 shows good agreement; the results are not affected qualitatively. The dominant contribution to the observed deviations originates from the distribution of mean activities over neurons within each population due to differences in synaptic drive. This affects the population-averaged autocorrelations a_α no longer agreeing to $m_\alpha(1-m_\alpha)$ in this case. An extension to distributed in-degrees can be done in the same manner as in the established mean-field theory [52].

One peculiarity arising in networks with a hard threshold is the independence of correlations of the synaptic coupling strength. This effect can be fully explained by the linearization procedure, because the susceptibility S of a neuron decreases inversely proportional to the strength of the synaptic noise σ. As the noise scales linearly in the synaptic amplitude, the effective linearized synaptic weight SJ is hence independent of J [63]. This also shows that a particular scaling of synaptic amplitudes with network size is not required to obtain small correlations in the balanced regime with irregular activity.

The presented work is closely related to our previous work on the correlation structure in spiking neuronal networks [17] and indeed was triggered by the review process of the latter. In [17], we exclusively

studied the symmetric connectivity structure, where excitatory and inhibitory neurons receive the same input on average. The results are qualitatively the same as those shown in Figure 4. A difference though is, that the external input in [17] is uncorrelated, whereas here it originates from a common finite population. The cancellation condition for input correlations, also observed in vivo [48], holds for spiking networks as well as for the binary networks studied here. The negative feedback suppressing fluctuations on the population level is the common underlying cause in both models. Future work needs to establish the formal relationship between the two network models.

Our theory presents a step towards an understanding of how correlated neuronal activity in local cortical circuits is shaped by recurrence and inputs from other cortical and thalamic areas. The correlation between membrane potentials of pairs of neurons in somatosensory cortex of behaving mice is dominated by low-frequency oscillations during quiet wakefulness. If the animal starts whisking, these correlations significantly decrease, even if the sensory nerve fibers are cut, suggesting an internal change of brain state [5]. Our work suggests that such a dynamic reduction of correlation could come about by modulating the effective negative feedback in the network. A possible neural implementation is the increase of tonic drive to inhibitory interneurons. This hypothesis is in line with the observed faster fluctuations in the whisking state [5]. Further work is needed to verify if such a mechanism yields a quantitative explanation of the experimental observations.

The network where the number of incoming external connections per neuron equals the size of the external population, cf. Figure 3 $N_x = K$, can be regarded as a setting where all neurons receive an identical incoming stimulus. The correlations between this signal and the responses of neurons in the local network (Figure 3C) are smaller than in an unconnected population without local negative feedback. This can formally be seen from (24), because negative eigenvalues of the recurrent coupling dampen the population response of the system. This suppression of correlations between stimulus and local activity hence implies weaker responses of single neurons to the driving signal. Recent experiments have shown that only a sparse subset of around 10 percent of the neurons in S1 of behaving mice responds to a sensory stimulus evoked by the active touch of a whisker with an object [4]. The subset of responding cells is determined by those neurons in which the cell specific combination of activated excitatory and inhibitory conductances drives the membrane potential above threshold. Our work suggests that negative feedback mediated among the layer 2/3 pyramidal cells, e.g. through local interneurons, should effectively reduce their correlated firing. In a biological network the negative feedback arrives with a synaptic delay and effectively reduces the low frequency content [17]. The response of the local activity is therefore expected to depend on the spectral properties of the stimulus. Intuitively one expects responses to better lock to the stimulus for fast and narrow transients with high frequency content. Further work is required to investigate this issue in more detail.

A large number of previous studies on the dynamics of local cortical networks focuses on the effect of the local connectivity, but ignores the spatio-temporal structure of external inputs by assuming that neurons in the local network are independently driven by external (often Poissonian) sources. Our study shows that the input correlations of pairs of neurons in the local network are only weakly affected by additional correlations caused by shared external afferents: Even for the extreme case where all neurons in the network receive exactly identical external input ($N_x = K$), the input correlations are small and only slightly larger than those obtained for the case where neurons receive uncorrelated external input ($N_x = 2N$; black curve in Figure 6C). One may therefore conclude that the approximation of uncorrelated external input is justified. In general, this may however be a hasty conclusion. Tiny changes in synaptic-input correlations have drastic effects, for example, on the power and reach of extracellular potentials [64]. For the modeling of extracellular potentials, knowledge of the spatio-temporal structure of inputs from remote areas is crucial.

The theory of correlations in presence of externally impinging signals is a required building block to study correlation-sensitive synaptic plasticity [65] in recurrent networks. Understanding the emerging structure of correlations imposed by an external signal is the first step in predicting the connectivity patterns resulting from ongoing synaptic plasticity sensitive to those correlations.

Acknowledgements

This work is partially supported by the Helmholtz Association: HASB and portfolio theme SMHB, the Next-Generation Supercomputer Project of MEXT, and EU grant 269921 (BrainScaleS). All simulations were carried out with NEST (http://www.nest-initiative.org).

References

1. Kilavik BE, Roux S, Ponce-Alvarez A, Confais J, Gruen S, et al. (2009) Long-term modifications in motor cortical dynamics induced by intensive practice. J Neurosci 29: 12653–12663.
2. Maldonado P, Babul C, Singer W, Rodriguez E, Berger D, et al. (2008) Synchronization of neuronal responses in primary visual cortex of monkeys viewing natural images. J Neurophysiol 100: 1523-1532.
3. Ito J, Maldonado P, Singer W, Grün S (2011) Saccade-related modulations of neuronal excitability support synchrony of visually elicited spikes. Cereb Cortex 21: 2482–2497.
4. Crochet S, Poulet JF, Kremer Y, Petersen CC (2011) Synaptic mechanisms underlying sparse coding of active touch. Neuron 69: 1160–1175.
5. Poulet J, Petersen C (2008) Internal brain state regulates membrane potential synchrony in barrel cortex of behaving mice. Nature 454: 881–885.
6. Salinas E, Sejnowski TJ (2001) Correlated neuronal activity and the flow of neural information. Nat Rev Neurosci 2: 539–550.
7. Abeles M (1982) Local Cortical Circuits: An Electrophysiological Study. Studies of Brain Function. Berlin, Heidelberg, New York: Springer-Verlag.
8. Diesmann M, Gewaltig MO, Aertsen A (1999) Stable propagation of synchronous spiking in cortical neural networks. Nature 402: 529–533.
9. Izhikevich EM (2006) Polychronization: Computation with spikes. Neural Comput 18: 245–282.
10. Sterne P (2012) Information recall using relative spike timing in a spiking neural network. Neural Comput 24: 2053–2077.
11. Hebb DO (1949) The organization of behavior: A neuropsychological theory. New York: John Wiley & Sons.
12. von der Malsburg C (1981) The correlation theory of brain function. Internal report 81-2, Department of Neurobiology, Max-Planck-Institute for Biophysical Chemistry, Göttingen, Germany.
13. Bienenstock E (1995) A model of neocortex. Network: Comput Neural Systems 6: 179–224.
14. Singer W, Gray C (1995) Visual feature integration and the temporal correlation hypothesis. Annu Rev Neurosci 18: 555–586.
15. Tripp B, Eliasmith C (2007) Neural populations can induce reliable postsynaptic currents without observable spike rate changes or precise spike timing. Cereb Cortex 17: 1830–1840.
16. Zohary E, Shadlen MN, Newsome WT (1994) Correlated neuronal discharge rate and its implications for psychophysical performance. Nature 370: 140–143.
17. Tetzlaff T, Helias M, Einevoll G, Diesmann M (2012) Decorrelation of neural-network activity by inhibitory feedback. PLoS Comput Biol 8: e1002596.
18. De la Rocha J, Doiron B, Shea-Brown E, Kresimir J, Reyes A (2007) Correlation between neural spike trains increases with firing rate. Nature 448: 802–807.

19. Rosenbaum R, Josic K (2011) Mechanisms that modulate the transfer of spiking correlations. Neural Comput 23: 1261–1305.

20. Rosenbaum R, Rubin JE, Doiron B (2013) Short-term synaptic depression and stochastic vesicle dynamics reduce and shape neuronal correlations. J Neurophysiol 109: 475–484.

21. Hertz J (2010) Cross-correlations in high-conductance states of a model cortical network. Neural Comput 22: 427–447.

22. Renart A, De La Rocha J, Bartho P, Hollender L, Parga N, et al. (2010) The asynchronous state in cortical cicuits. Science 327: 587–590.

23. Shadlen MN, Newsome WT (1998) The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding. J Neurosci 18: 3870–3896.

24. Tetzlaff T, Rotter S, Stark E, Abeles M, Aertsen A, et al. (2008) Dependence of neuronal correlations on filter characteristics and marginal spike-train statistics. Neural Comput 20: 2133–2184.

25. Kriener B, Tetzlaff T, Aertsen A, Diesmann M, Rotter S (2008) Correlations and population dynamics in cortical networks. Neural Comput 20: 2185–2226.

26. Pernice V, Staude B, Cardanobile S, Rotter S (2011) How structure determines correlations in neuronal networks. PLoS Comput Biol 7: e1002059.

27. Trousdale J, Hu Y, Shea-Brown E, Josic K (2012) Impact of network structure and cellular response on spike time correlations. PLoS Comput Biol 8: e1002408.

28. Helias M, Tetzlaff T, Diesmann M (2013) Echoes in correlated neural systems. New J Phys 15: 023002.

29. Pernice V, Staude B, Cardanobile S, Rotter S (2012) Recurrent interactions in spiking networks with arbitrary topology. Phys Rev E 85: 031916.

30. Bi G, Poo M (1998) Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type. J Neurosci 18: 10464–10472.

31. Gilson M, Burkitt AN, Grayden DB, Thomas DA, van Hemmen JL (2009) Emergence of network structure due to spike-timing-dependent plasticity in recurrent neuronal networks. I. Input selectivity - strengthening correlated input pathways. Biol Cybern 101: 81–102.

32. Lindén H, Tetzlaff T, Potjans TC, Pettersen KH, Grün S, et al. (2011) Modeling the spatial reach of the LFP. Neuron 72: 859–872.

33. Ginzburg I, Sompolinsky H (1994) Theory of correlations in stochastic neural networks. Phys Rev E 50: 3171–3191.

34. Meyer C, van Vreeswijk C (2002) Temporal correlations in stochastic networks of spiking neurons. Neural Comput 14: 369–404.

35. Lindner B, Doiron B, Longtin A (2005) Theory of oscillatory firing induced by spatially correlated noise and delayed inhibitory feedback. Phys Rev E 72: 061919.

36. Ostojic S, Brunel N, Hakim V (2009) How connectivity, background activity, and synaptic properties shape the cross-correlation between spike trains. J Neurosci 29: 10234–10253.

37. Hu Y, Trousdale J, Josić K, Shea-Brown E (2013) Motif statistics and spike correlations in neuronal networks. J Stat Mech : P03012.

38. Brunel N, Hakim V (1999) Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. Neural Comput 11: 1621–1671.

39. Litwin-Kumar A, Chacron MJ, Doiron B (2012) The spatial structure of stimuli shapes the timescale of correlations in population spiking activity. PLoS Comput Biol 8: e1002667.

40. van Vreeswijk C, Sompolinsky H (1995) Theory of randomly connected networks with excitation-inhibition balance. Cortical Dynamics in Jerusalem (Program and Abstracts) : 73.

41. Amit DJ, Brunel N (1997) Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. Cereb Cortex 7: 237–252.

42. Brunel N (2000) Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. J Comput Neurosci 8: 183–208.

43. Potjans TC, Diesmann M (2012) The cell-type specific cortical microcircuit: Relating structure and activity in a full-scale spiking network model. Cerebral Cortex : doi: 10.1093/cercor/bhs358.

44. Binzegger T, Douglas RJ, Martin KAC (2004) A quantitative map of the circuit of cat primary visual cortex. J Neurosci 39: 8441–8453.

45. Stepanyants A, Martinez LM, Ferecskó AS, Kisvárday ZF (2009) The fractions of short- and long-range connections in the visual cortex. Proc Nat Acad Sci USA 106: 3555–3560.

46. Gilbert CD, Wiesel TN (1983) Clustered intrinsic connections in cat visual cortex. J Neurosci 5: 1116–33.

47. Voges N, Schüz A, Aertsen A, Rotter S (2010) A modeler's view on the spatial structure of intrinsic horizontal connectivity in the neocortex. Progress in Neurobiology 92: 277–292.

48. Okun M, Lampl I (2008) Instantaneous correlation of excitation and inhibition during sensory-evoked activities. Nat Neurosci 11: 535–537.

49. Gentet L, Avermann M, Matyas F, Staiger JF, Petersen CC (2010) Membrane potential dynamics of GABAergic neurons in the barrel cortex of behaving mice. Neuron 65: 422–435.

50. van Vreeswijk C, Sompolinsky H (1996) Chaos in neuronal networks with balanced excitatory and inhibitory activity. Science 274: 1724–1726.

51. Parga N (2013) Towards a self-consistent description of irregular and asynchronous cortical activity. J Stat Mech: Theory and Exp : P03010.

52. van Vreeswijk C, Sompolinsky H (1998) Chaotic balanced state in a model of cortical circuits. Neural Comput 10: 1321–1371.

53. Buice MA, Cowan JD, Chow CC (2009) Systematic fluctuation expansion for neural network activity equations. Neural Comput 22: 377–426.

54. Rumelhart DE, McClelland JL, the PDP Research Group (1986) Parallel Distributed Processing, Explorations in the Microstructure of Cognition: Foundations, volume 1. Cambridge, Massachusetts: MIT Press.

55. Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci USA 79: 2554–2558.

56. Hanuschkin A, Kunkel S, Helias M, Morrison A, Diesmann M (2010) A general and efficient method for incorporating precise spike times in globally time-driven simulations. Front Neuroinform 4: 113.

57. Gewaltig MO, Diesmann M (2007) NEST (NEural Simulation Tool). Scholarpedia 2: 1430.

58. Hertz J, Krogh A, Palmer RG (1991) Introduction to the Theory of Neural Computation. Perseus Books.

59. Kelly F (1979) Stochastic processes and reversibility. Wiley, Cambridge University Press.

60. Jones E, Oliphant T, Peterson P, et al. (2001). SciPy: Open source scientific tools for Python. Http://www.scipy.org/.

61. Tetzlaff T, Helias M, Einevoll GT, Diesmann M (2012) Decorrelation of neural-network activity by inhibitory feedback. arXiv : 1204.4393v1 [q-bio.NC].

62. Buzsáki G, Wang XJ (2012) Mechanisms of gamma oscillations. Annu Rev Neurosci 35: 203–225.

63. Grytskyy D, Tetzlaff T, Diesmann M, Helias M (2013) Invariance of covariances arises out of noise. AIP Conf Proc 1510: 258–262.

64. Lindén H, Tetzlaff T, Potjans TC, Pettersen KH, Grün S, et al. (2011) Modeling the spatial reach of the LFP. Neuron 72: 859–872.

65. Morrison A, Diesmann M, Gerstner W (2008) Phenomenological models of synaptic plasticity based on spike-timing. Biol Cybern 98: 459–478.